Prediction of whole-genome DNA–DNA similarity, determination of G + C content and phylogenetic analysis within the family *Pasteurellaceae* by multilocus sequence analysis (MLSA)

Peter Kuhnert and Bożena M. Korczak

Correspondence

Peter Kuhnert
peter.kuhnert@vbi.unibe.ch

Institute of Veterinary Bacteriology, Vetsuisse Faculty of the University of Bern, Laenggass-Str. 122, CH-3001 Bern, Switzerland

Genome predictions based on selected genes would be a very welcome approach for taxonomic studies, including DNA–DNA similarity, G + C content and representative phylogeny of bacteria. At present, DNA–DNA hybridizations are still considered the gold standard in species descriptions. However, this method is time-consuming and troublesome, and datasets can vary significantly between experiments as well as between laboratories. For the same reasons, full matrix hybridizations are rarely performed, weakening the significance of the results obtained. The authors established a universal sequencing approach for the three genes *recN*, *rpoA* and *thdF* for the *Pasteurellaceae*, and determined if the sequences could be used for predicting DNA–DNA relatedness within the family. The sequence-based similarity values calculated using a previously published formula proved most useful for species and genus separation, indicating that this method provides better resolution and no experimental variation compared to hybridization. By this method, cross-comparisons within the family over species and genus borders easily become possible. The three genes also serve as an indicator of the genome G + C content of a species. A mean divergence of around 1 % was observed from the classical method, which in itself has poor reproducibility. Finally, the three genes can be used alone or in combination with already-established 16S rRNA, *rpoB* and *infB* gene-sequencing strategies in a multisequence-based phylogeny for the family *Pasteurellaceae*. It is proposed to use the three sequences as a taxonomic tool, replacing DNA–DNA hybridization.

Received 13 March 2006

Revised 19 May 2006

Accepted 23 May 2006

INTRODUCTION

The family *Pasteurellaceae* (Pohl, 1981) contains bacterial species that colonize the mucous membranes of the alimentary, genital or respiratory tract of vertebrates, including man and other mammals, birds and reptiles. Most of these bacteria are commensals, but a few are important pathogens, such as: *Haemophilus influenzae*, causing otitis media and meningitis in man; *Pasteurella multocida*, the aetiological agent of fowl cholera, haemorrhagic septicaemia in cattle and water buffaloes, and atrophic rhinitis in swine; *Mannheimia haemolytica*, causing ‘shipping fever’ in cattle; and *Actinobacillus pleuropneumoniae*, responsible for porcine pleuropneumonia (Donachie *et al.*, 1995). For a long time, the family only included the old genera *Pasteurella*, *Actinobacillus* and *Haemophilus*. Currently, it contains the additional eight genera

Mannheimia, *Histophilus*, *Phocoenobacter*, *Lonepinella*, *Gallibacterium*, *Volucribacter*, *Avibacterium* and *Nicoletella*. More than 60 validated species are members of the family (www.the-icsp.org/taxa/Pasteurellaceaealist.htm). Many more species, as well as genera, are expected to be described in the near future.

The taxonomy of the *Pasteurellaceae* is complex and at present incompletely resolved. A number of organisms, mainly (but unfortunately not only) those classified in the past, need to be reclassified. Moreover, many taxa await final description and classification, e.g. several genomospecies and numerous Bisgaard taxa (Christensen *et al.*, 2003b). Based on monophyletic clusters observed with 16S rRNA gene (*rrs*) phylogeny, a minimum of 30 genera are expected for the family. Due to many closely related taxa, phenotypic identification and separation are problematic in certain cases. The requirement for phenotypic characteristics that separate new species from related taxa is often an obstacle for new species descriptions, leaving diagnostic laboratories with unnamed genomospecies. Further complicating a precise classification of new taxa is the need for DNA–DNA

Abbreviations: MLSA, multilocus sequence analysis; MLSP, multilocus sequence phylogeny.

Individual phylogenetic trees for six genes are available as supplementary data with the online version of this paper.

hybridization, which is still regarded as the gold standard for species definition and description. However, this technique is cumbersome, if not impossible, to perform with some taxa. In addition, the method shows a very high variation between experiments and between laboratories. Due to these difficulties and their labour-intensive nature, full matrix hybridizations are rarely carried out, which weakens the significance of the results obtained. Moreover, for each new taxon suggested, cross-hybridization with all the relevant taxa is needed, and the use of reference strains is a prerequisite. Finally, the method only allows investigation at the species level, since its use for investigating genera is less trustworthy, due to its technical limitations and low resolution. Therefore, the 'ad hoc committee for the re-evaluation of species definition in bacteriology' has suggested the development and validation of alternative techniques to DNA–DNA hybridization (Stackebrandt *et al.*, 2002). However, phylogenetic analysis by *rrs* sequences, used successfully for years, has its limitations for classification. It does not allow analysis of closely related taxa, and in certain cases contradicts other classification markers. The exploitation of alternative housekeeping genes and their use in multilocus sequence analysis (MLSA) should therefore be investigated (Gevers *et al.*, 2005). Several studies using housekeeping genes have been done within the *Pasteurellaceae*. They include the use of *infB* for the genera *Haemophilus* and *Actinobacillus* (Hedegaard *et al.*, 2001; Norskov-Lauritsen *et al.*, 2004), the *rpoB* gene for the entire family (Korczak *et al.*, 2004), and *sodA* for the genus *Pasteurella* (Gautier *et al.*, 2005). Investigations have also included comparisons of *rrs*, *rpoB*, *infB* and *atpD* (Christensen *et al.*, 2004b). These studies have shown that these genes represent useful phylogenetic markers which can be applied in taxonomy, but they are not suitable for the prediction of whole-genome relatedness as it is assessed by DNA–DNA hybridization. Recently, Zeigler (2003) has shown that a selection of three genes can be used to predict whole-genome relatedness. Comparing whole-genome sequences of 49 bacterial species, he concluded that out of 32 candidate genes investigated, similarities of the genes *recN*, *thdF* and *rpoA*, used in a formula, can be applied to predict confidently the whole-genome similarities among the respective taxa. Based on these findings, we developed a sequencing strategy for *recN* (encoding a DNA repair protein), *thdF* (encoding a GTPase) and *rpoA* (encoding the alpha subunit of the RNA polymerase) within the *Pasteurellaceae*, and validated their use in MLSA to predict the whole-genome DNA–DNA similarity, G+C content and phylogeny of selected taxa.

METHODS

Amplification and sequencing of housekeeping genes.

Amplification was carried out in 50 µl volumes containing 20 pmol each primer, 1 mM dNTP, 1× reaction buffer B (supplied with FIREPol DNA polymerase), 2.5 mM MgCl₂ and 2.5 U FIREPol polymerase (Solis BioDyne). Approximately 100 ng template was added as genomic DNA or as lysate. Cycling conditions were 3 min denaturation at 94 °C, followed by 35 cycles at 94 °C for 30 s, 54 °C for 30 s and 72 °C for 1 min. A final extension step for 7 min at

72 °C was included. PCR products were purified with the High Pure PCR Purification kit (Roche Applied Science). In cases where non-specific bands were observed, the specific band was purified from agarose gels. Finally, about 30 ng purified PCR product was used for sequencing with the BigDye Terminator cycle sequencing kit (Applied Biosystems). Sequences were analysed on an ABI Prism 3100 Genetic Analyser (Applied Biosystems) and then edited using the Sequencer software (GeneCodes).

Sequence determination of *rrs*, *rpoB* and *infB* genes of *Pasteurellaceae* was done as previously published (Kuhnert *et al.*, 2002, 2004; Korczak *et al.*, 2004).

Sequence analysis. Sequence similarities of the three genes from different species were calculated using the CLUSTALX program generating a distance matrix (Thompson *et al.*, 1997). Values were converted to a similarity matrix in Microsoft Excel and then used in the formula described by Zeigler (2003):

$$SI_{\text{genome}} = -1.88 + 0.52(SI_{\text{recN}}) + 1.78(SI_{\text{thdF}}) + 0.52(SI_{\text{rpoA}}) \text{ or}$$

$$SI_{\text{genome}} = -1.30 + 2.25(SI_{\text{itrecN}})$$

where SI is sequence identity. The mol% G+C content was calculated from the base composition of the three genes. Phylogenetic analysis was carried out using Bionumerics version 4.0 (Applied Maths).

RESULTS AND DISCUSSION

Primer selection and sequence determination

Based on available genome sequences of *Pasteurellaceae* species, consensus primers were selected for *rpoA*, *recN* and *thdF*. First of all, they had to be as universal as possible for the entire family in order to minimize numbers of primers used for sequence determination. Secondly, in order to obtain optimal information for the calculation of genome relatedness, primers located as close as possible to the 5' and 3' ends of the genes had to be chosen. Finally, the same primers had to be used for PCR and sequencing in order to obtain a simple and straightforward approach. A list of primers used, and their sequences and location, is given in Table 1. With these primers, almost the full length of the *rpoA* (1 kb) and *thdF* (1.4 kb) genes, as well as 1.4 kb of the 1.7 kb *recN* gene, could be obtained. PCR worked very well for all species with the *rpoA* gene, which is the most conserved of the three genes used. With the less-conserved *thdF* and *recN* genes, multiple bands were observed, mainly in the genus *Actinobacillus*. In these cases, internal sequencing primers were employed, or specific bands were excised and purified from agarose gels. For certain species, optimization of PCR and sequencing primers was necessary. Compared to the problems associated with conventional hybridization, a one-off selection of additional primers and optimization of temperatures for certain taxa represents only a minor problem.

For *recN*, in most cases a mixture of the four primers *recN*-L, *recN*-L2, *recN*-R and *recN*-R2 was used for PCR and for sequencing. Exceptions to this were [*Actinobacillus*] *indolicus* and [*Haemophilus*] *parasuis*, for which *recN*-L2/*recN*-R2 alone gave a PCR product. Primer *recN*_Vp-R was used

Table 1. Primers used for amplification and sequencing

Name	Sequence (5'→3')	Position*
rpoA-L	TCTGTRACAGAATTTTTAAARCC	10–32
rpoA-R	TTGCNNGGNGGCCARTTTTCAAGG	973–951
recN-L	CAACTYACTATYAATMATTTTGC	10–32
recN-R	CTAATGCCYACRTCYACTTCATC	1397–1375
recN-L2	CATTTAACGGTTAATAATTTTGC	10–32
recN-R2	CTAATYCCMACATCNACYTCATC	1397–1375
recN_Vp-R	CTGATCCCTACATCAATTTTCATC	1397–1375
recN_Ga-1	GCTATTTAGATGAATTGGTTGAGG	752–775
recN_Ga-2	CCTCAACCAATTCATCTAAATAGC	775–752
thdF_first-L2	ATGAAAGANACNATYGTNGCWCARGC	7–32
thdF_first-R2	TGTTTATTTNCCRATRCARAANGARC	1368–1343
thdF-L2	AAAGANACNATYGTNGCWCARGC	10–32
thdF-R2	TTATTTNCCRATRCARAANGARC	1365–1343
thdF_MP-L	AAAGAMACCATTGTTGCWCAAGC	10–32
thdF-1	GGAATGAAAGTCGTGATTGCAGG	646–668
thdF-2	CCTGCAATCACGACTTTCATTCC	668–646
thdF-3	GGATGAAAGTGGTAATTGC	647–665
thdF-4	GCAATTACCACTTTCATCC	665–647

*In relation to the respective *E. coli* sequence.

instead of the two R primers (recN-R and recN-R2) for *Volucrobacter psittacida*, *Avibacterium gallinarum* und *Mannheimia ruminalis*. For *Gallibacterium anatis*, internal sequencing primers recN_Ga-1 and recN_Ga-2 were necessary to obtain the full sequence.

For *thdF*, the primers thdF_first-L2 and thdF_first-R2 were generally used as PCR and sequencing primers. Primers thdF-L2 and thdF-R2 were used as alternative primers during the study, although unlike with *recN*, never in combination with the other two. Primers thdF-1 and thdF-2 were used as internal sequencing primers for all strains, except for *Actinobacillus hominis*, *Actinobacillus suis*, *Actinobacillus equuli* subsp. *haemolyticus*, *Actinobacillus arthritidis*, *Actinobacillus* genomospecies 1 and *Pasteurella multocida* subsp. *gallicida*, for which internal primers thdF-3 and thdF-4 were used. For *A. gallinarum* and [*Actinobacillus*] *minor*, primer thdF_MP-L was used instead of thdF_first-L2 for PCR and sequencing.

The standard annealing temperature for PCR was 54 °C, but in certain cases this had to be lowered to 48 °C (e.g. *recN* with *G. anatis*) or increased to 58 °C (e.g. *thdF* with *A. pleuropneumoniae* and *Mannheimia* species).

Genome relatedness of the genera

The three genes *rpoA*, *recN* and *thdF* of 43 strains (37 species and subspecies, including all type species of named genera) from the family *Pasteurellaceae* were amplified and sequenced. A list of all the strains included and the accession numbers of the sequences are given in Table 2. Based on the obtained sequences, whole-genome relatedness between the

species was calculated and is given as similarity values in Table 3. For the first time, cross-comparisons among species and genera were possible. With the classical hybridization method of Brenner *et al.* (1982), the DNA–DNA relatedness at species level was set at 70 % (Wayne *et al.*, 1987). Within the family *Pasteurellaceae*, data have mainly been generated with the spectrophotometric method, and Mutters *et al.* (1989) have shown that a limit of 85 % for species separation is reasonable. Moreover, a homology above 55 % is set for genus separation with the *Pasteurellaceae*. However, the use of DNA–DNA hybridization to investigate less-related organisms (e.g. two different genera) does not provide reproducible results at this level. Application of the presented sequence method has no limitations, since similarity values are calculated and are not dependent on experimental settings. Therefore, genomes of less-related species can be compared and the calculated values can be used for taxonomic purposes. Based upon the data in the lower left of Table 3, preliminary threshold values can be set at around 0.85 for species separation, while the limit for genus separation is about 0.4, i.e. strains showing similarity values below 0.85 almost certainly belong to two different species, whereas strains showing similarity values below 0.4 most likely belong to two different genera. Strains of the same species show similarity values of 0.9 and higher, and species of the same genus have similarity values above 0.4. Similarity values between 0.85 and 0.9 are intermediate, and might indicate different species or subspecies, as observed for the two species *M. haemolytica*/*Mannheimia glucosida* (0.88) or with the two subspecies of *A. equuli* (0.87). These limits for species and genus delineation are preliminary and generalized suggestions, and might change as new data

Table 2. Strains used in this study and accession numbers of genes

No.	Species	Accession number		
		<i>recN</i>	<i>rpoA</i>	<i>thdF</i>
1	<i>Pasteurella multocida</i> CCUG 17976 ^T	DQ410890	DQ410928	DQ410966
2	<i>Haemophilus influenzae</i> Rd	NC_000907	NC_000907	NC_000907
3	<i>Histophilus somni</i> HS8025 ^T	DQ410891	DQ410929	DQ410967
4	<i>Gallibacterium anatis</i> CCUG 15563 ^T	DQ410892	DQ410930	DQ410968
5	<i>Volucrobacter psittacidia</i> JEO101	DQ410893	DQ410931	DQ410969
6	<i>Avibacterium gallinarum</i> NCTC 11188 ^T	DQ410894	DQ410932	DQ410970
7	<i>Nicotella semolina</i> CCUG 43639 ^T	DQ410895	DQ410933	DQ410971
8	<i>Phocobacter uteri</i> NCTC 12872 ^T	DQ410896	DQ410934	DQ410972
9	<i>Lonepinella koalarum</i> ATCC 700131 ^T	DQ410897	DQ410935	DQ410973
10	<i>Mannheimia haemolytica</i> NCTC 9380 ^T	DQ410898	DQ410936	DQ410974
11	[<i>Haemophilus</i>] <i>ducreyi</i> 35000HP	AE017143	AE017143	AE017143
12	[<i>Pasteurella</i>] <i>trehalosi</i> NCTC 10370 ^T	DQ410899	DQ410937	DQ410975
13	[<i>Actinobacillus</i>] <i>actinomycetemcomitans</i> HK1651	www.genome.ou.edu		
14	[<i>Mannheimia</i>] <i>succiniciproducens</i> MBEL55E	NC_006300	NC_006300	NC_006300
15	[<i>Actinobacillus</i>] <i>minor</i> NM305 ^T	DQ410900	DQ410938	DQ410976
16	[<i>Actinobacillus</i>] <i>porcitonsillarum</i> CCUG 46996	DQ410901	DQ410939	DQ410977
17	<i>Actinobacillus</i> genomospecies 2 CCUG 15571	DQ410902	DQ410940	DQ410978
18	<i>Actinobacillus hominis</i> CCUG 19800 ^T	DQ410903	DQ410941	DQ410979
19	<i>Actinobacillus ureae</i> CCUG 2139 ^T	DQ410904	DQ410942	DQ410980
20	<i>Actinobacillus arthritidis</i> CCUG 24862 ^T	DQ410905	DQ410943	DQ410981
21	<i>Actinobacillus suis</i> ATCC 33415 ^T	DQ410906	DQ410944	DQ410982
22	<i>Actinobacillus</i> genomospecies 1 CCUG 22229	DQ410907	DQ410945	DQ410983
23	<i>Actinobacillus capsulatus</i> CCUG 12396 ^T	DQ410908	DQ410946	DQ410984
24	<i>Actinobacillus equuli</i> subsp. <i>equuli</i> ATCC 19392 ^T	DQ410909	DQ410947	DQ410985
25	<i>A. equuli</i> subsp. <i>haemolyticus</i> CCUG 19799 ^T	DQ410910	DQ410948	DQ410986
26	<i>Actinobacillus lignieresii</i> NCTC 4189 ^T	DQ410911	DQ410949	DQ410987
27	<i>Actinobacillus pleuropneumoniae</i> serotype 1 S4074 ^T	DQ410912	DQ410950	DQ410988
28	<i>A. pleuropneumoniae</i> serotype 2 S1536	DQ410913	DQ410951	DQ410989
29	<i>A. pleuropneumoniae</i> serotype 3 S1421	DQ410914	DQ410952	DQ410990
30	<i>A. pleuropneumoniae</i> serotype 4 M62	DQ410915	DQ410953	DQ410991
31	<i>A. pleuropneumoniae</i> serotype 6 FemØ	DQ410916	DQ410954	DQ410992
32	<i>A. pleuropneumoniae</i> serotype 8 405	DQ410917	DQ410955	DQ410993
33	[<i>Actinobacillus</i>] <i>indolicus</i> 46KC2 ^T	DQ410918	DQ410956	DQ410994
34	[<i>Actinobacillus</i>] <i>porcinus</i> NM319 ^T	DQ410919	DQ410957	DQ410995
35	[<i>Haemophilus</i>] <i>parasuis</i> CCUG 3712 ^T	DQ410920	DQ410958	DQ410996
36	<i>Mannheimia glucosida</i> CCUG 38457 ^T	DQ410921	DQ410959	DQ410997
37	<i>Mannheimia varigena</i> CCUG 38462 ^T	DQ410922	DQ410960	DQ410998
38	<i>Mannheimia ruminalis</i> CCUG 38470 ^T	DQ410923	DQ410961	DQ410999
39	<i>Mannheimia granulomatis</i> ATCC 49244 ^T	DQ410924	DQ410962	DQ411000
40	<i>Pasteurella multocida</i> subsp. <i>multocida</i> Pm70	AE004439	AE004439	AE004439
41	<i>P. multocida</i> subsp. <i>gallicida</i> CCUG 17978 ^T	DQ410925	DQ410963	DQ411001
42	<i>P. multocida</i> subsp. <i>septica</i> CCUG 17977 ^T	DQ410926	DQ410964	DQ411002
43	<i>Pasteurella canis</i> CCUG 12400 ^T	DQ410927	DQ410965	DQ411003

accumulate, or might have to be adapted for certain groups within the family. As DNA–DNA similarity is used as a part of a polyphasic approach, criteria other than sequence-based prediction of genome similarity should be included for proper classification of such strains. The highest similarity value that can be obtained with 100 % identical sequences in the formula is 0.94. The fact that this value is not 1.0

somehow reflects the possibility of mobile genetic elements changing the overall similarity of genomes. Phages, plasmids and integrons can change a single strain and lead to strain variation, even though the rest of the genome is identical. Thereby, the mathematical model respects the biological fact of genome plasticity within species and even within strains (D. R. Zeigler, personal communication).

Table 3. Whole-genome DNA–DNA similarity value matrix based on calculation using the three genes *recN*, *rpoA* and *thdF* (lower left) or *recN* alone (upper right)

No.	Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	<i>P. multocida</i>		0.199	0.178	0.012	0.228	0.176	0.005	0.061	0.246	0.052	0.048	0.061	0.192	0.183	0.057	0.061	0.055	0.070	0.077	0.045	0.052
2	<i>H. influenzae</i>	0.262		0.212	−0.004	0.136	0.190	0.048	0.088	0.232	0.111	0.106	0.118	0.203	0.250	0.091	0.077	0.081	0.057	0.059	0.075	0.066
3	<i>H. somni</i>	0.220	0.209		0.021	0.106	0.190	−0.033	0.007	0.187	0.034	0.007	0.063	0.136	0.158	0.007	−0.011	0.032	0.016	0.023	0.019	0.034
4	<i>G. anatis</i>	0.167	0.095	0.134		−0.002	−0.002	−0.040	−0.018	−0.029	−0.027	0.012	−0.004	−0.018	0.016	−0.033	−0.031	0.027	0.025	0.025	0.012	0.023
5	<i>V. psittacida</i>	0.272	0.207	0.229	0.142		0.075	0.068	0.088	0.163	0.084	0.091	0.079	0.115	0.156	0.097	0.099	0.091	0.097	0.102	0.075	0.095
6	<i>A. gallinarum</i>	0.256	0.227	0.260	0.125	0.174		0.037	0.066	0.226	0.070	0.032	0.073	0.151	0.185	0.057	0.052	0.027	0.021	0.030	0.014	0.034
7	<i>N. semolina</i>	0.102	0.088	0.067	0.056	0.123	0.134		0.246	0.021	0.253	0.282	0.309	0.023	0.057	0.289	0.259	0.311	0.322	0.327	0.307	0.322
8	<i>P. uteri</i>	0.162	0.122	0.140	0.094	0.149	0.181	0.189		0.077	0.244	0.255	0.268	0.032	0.045	0.268	0.284	0.264	0.264	0.275	0.262	0.266
9	<i>L. koalarum</i>	0.273	0.311	0.237	0.149	0.235	0.281	0.095	0.187		0.070	0.075	0.061	0.185	0.208	0.070	0.073	0.048	0.039	0.045	0.043	0.070
10	<i>M. haemolytica</i>	0.143	0.136	0.105	0.127	0.129	0.175	0.231	0.243	0.158		0.275	0.298	0.027	0.050	0.322	0.304	0.268	0.271	0.273	0.257	0.257
11	[<i>H.</i>] <i>ducreyi</i>	0.143	0.119	0.135	0.111	0.161	0.150	0.233	0.245	0.168	0.270		0.311	−0.002	0.061	0.320	0.309	0.338	0.338	0.336	0.331	0.356
12	[<i>P.</i>] <i>trehalosi</i>	0.136	0.156	0.132	0.118	0.114	0.154	0.248	0.246	0.149	0.318	0.275		0.052	0.061	0.291	0.291	0.327	0.334	0.334	0.325	0.320
13	[<i>A.</i>] <i>actinomycetemcomitans</i>	0.269	0.220	0.184	0.127	0.190	0.212	0.068	0.102	0.232	0.132	0.092	0.124		0.266	0.023	0.057	0.019	0.039	0.043	0.021	0.021
14	[<i>M.</i>] <i>succiniciproducens</i>	0.295	0.227	0.214	0.131	0.197	0.239	0.081	0.124	0.291	0.152	0.120	0.121	0.287		0.037	0.027	0.027	0.075	0.084	0.039	0.048
15	[<i>A.</i>] <i>minor</i>	0.172	0.154	0.133	0.109	0.129	0.172	0.251	0.261	0.151	0.317	0.329	0.306	0.146	0.155		0.786	0.302	0.322	0.320	0.302	0.318
16	[<i>A.</i>] <i>porcitonsillarum</i>	0.159	0.140	0.139	0.106	0.143	0.171	0.245	0.268	0.159	0.311	0.306	0.304	0.138	0.169	0.789		0.316	0.322	0.320	0.307	0.322
17	<i>Actinobacillus</i> genomospecies 2	0.165	0.143	0.135	0.104	0.149	0.164	0.245	0.269	0.161	0.329	0.374	0.313	0.155	0.183	0.337	0.347		0.649	0.649	0.912	0.667
18	<i>A. hominis</i>	0.152	0.124	0.116	0.098	0.150	0.166	0.238	0.251	0.169	0.324	0.370	0.311	0.169	0.196	0.355	0.349	0.681		0.948	0.646	0.673
19	<i>A. ureae</i>	0.156	0.117	0.123	0.104	0.145	0.167	0.218	0.268	0.164	0.328	0.357	0.305	0.158	0.189	0.339	0.345	0.713	0.819		0.646	0.676
20	<i>A. arthritidis</i>	0.167	0.129	0.127	0.116	0.154	0.175	0.240	0.286	0.162	0.334	0.376	0.303	0.166	0.181	0.330	0.338	0.848	0.696	0.724		0.678
21	<i>A. suis</i>	0.142	0.120	0.129	0.104	0.147	0.172	0.233	0.247	0.173	0.328	0.365	0.306	0.150	0.180	0.337	0.333	0.706	0.771	0.787	0.727	
22	<i>Actinobacillus</i> genomospecies 1	0.146	0.108	0.117	0.121	0.149	0.161	0.232	0.260	0.180	0.328	0.358	0.288	0.158	0.198	0.345	0.354	0.729	0.735	0.763	0.738	0.761
23	<i>A. capsulatus</i>	0.155	0.112	0.127	0.108	0.159	0.167	0.242	0.257	0.174	0.324	0.377	0.313	0.151	0.188	0.356	0.352	0.687	0.774	0.773	0.700	0.804
24	<i>A. equuli</i> subsp. <i>equuli</i>	0.146	0.109	0.122	0.100	0.152	0.158	0.227	0.244	0.152	0.329	0.359	0.301	0.144	0.180	0.346	0.343	0.710	0.770	0.808	0.704	0.800
25	<i>A. equuli</i> subsp. <i>haemolyticus</i>	0.148	0.115	0.127	0.108	0.154	0.151	0.222	0.246	0.161	0.334	0.371	0.310	0.155	0.186	0.345	0.354	0.723	0.768	0.815	0.729	0.807
26	<i>A. lignieresii</i>	0.142	0.118	0.120	0.099	0.136	0.164	0.244	0.246	0.163	0.312	0.346	0.290	0.155	0.167	0.329	0.329	0.691	0.715	0.716	0.714	0.798
27	<i>A. pleuropneumoniae</i> serotype 1	0.149	0.116	0.127	0.098	0.134	0.165	0.241	0.248	0.158	0.321	0.352	0.286	0.155	0.160	0.333	0.334	0.700	0.718	0.707	0.711	0.788
28	<i>A. pleuropneumoniae</i> serotype 2	0.144	0.113	0.125	0.093	0.132	0.159	0.242	0.242	0.160	0.316	0.350	0.286	0.150	0.156	0.330	0.333	0.697	0.720	0.710	0.712	0.795
29	<i>A. pleuropneumoniae</i> serotype 4	0.146	0.117	0.123	0.096	0.130	0.163	0.243	0.244	0.160	0.316	0.352	0.286	0.149	0.158	0.330	0.332	0.698	0.715	0.707	0.712	0.791
30	<i>A. pleuropneumoniae</i> serotype 4	0.147	0.114	0.126	0.097	0.130	0.162	0.243	0.244	0.161	0.319	0.351	0.289	0.150	0.161	0.329	0.331	0.699	0.720	0.711	0.714	0.800
31	<i>A. pleuropneumoniae</i> serotype 6	0.146	0.114	0.126	0.097	0.131	0.160	0.243	0.244	0.162	0.319	0.352	0.289	0.151	0.161	0.329	0.332	0.699	0.720	0.711	0.714	0.802
32	<i>A. pleuropneumoniae</i> serotype 8	0.149	0.115	0.127	0.096	0.133	0.160	0.243	0.245	0.163	0.319	0.349	0.288	0.149	0.160	0.329	0.334	0.698	0.716	0.706	0.714	0.791
33	[<i>A.</i>] <i>indolicus</i>	0.178	0.136	0.104	0.081	0.164	0.232	0.331	0.266	0.159	0.335	0.269	0.369	0.154	0.193	0.330	0.325	0.344	0.357	0.341	0.350	0.359
34	[<i>A.</i>] <i>porcinus</i>	0.265	0.273	0.264	0.172	0.249	0.300	0.126	0.197	0.364	0.213	0.168	0.168	0.290	0.315	0.205	0.205	0.217	0.218	0.216	0.199	0.213
35	[<i>H.</i>] <i>parasuis</i>	0.196	0.170	0.135	0.101	0.176	0.219	0.310	0.288	0.180	0.317	0.277	0.365	0.150	0.189	0.348	0.340	0.330	0.339	0.323	0.342	0.325
36	<i>M. glucosida</i>	0.145	0.138	0.113	0.119	0.129	0.180	0.235	0.238	0.164	0.881	0.264	0.316	0.141	0.160	0.319	0.313	0.336	0.336	0.343	0.348	0.342
37	<i>M. varigena</i>	0.143	0.101	0.089	0.105	0.114	0.184	0.220	0.240	0.167	0.511	0.262	0.301	0.129	0.156	0.321	0.310	0.326	0.339	0.333	0.337	0.357
38	<i>M. ruminalis</i>	0.155	0.131	0.114	0.120	0.134	0.166	0.227	0.248	0.161	0.744	0.281	0.318	0.123	0.157	0.327	0.331	0.351	0.335	0.345	0.356	0.348
39	<i>M. granulomatis</i>	0.144	0.113	0.105	0.087	0.125	0.173	0.243	0.243	0.139	0.482	0.274	0.295	0.112	0.129	0.301	0.305	0.367	0.344	0.344	0.374	0.339
40	<i>P. multocida</i> Pm70	0.910	0.264	0.215	0.167	0.266	0.252	0.109	0.154	0.266	0.143	0.135	0.135	0.257	0.292	0.171	0.157	0.159	0.148	0.147	0.156	0.134
41	<i>P. multocida</i> subsp. <i>gallicida</i>	0.922	0.260	0.215	0.161	0.271	0.253	0.106	0.163	0.266	0.143	0.136	0.136	0.266	0.299	0.173	0.161	0.167	0.155	0.159	0.169	0.143
42	<i>P. multocida</i> subsp. <i>septica</i>	0.883	0.250	0.221	0.171	0.278	0.249	0.107	0.153	0.261	0.152	0.137	0.137	0.272	0.292	0.170	0.157	0.161	0.142	0.151	0.158	0.140
43	<i>P. canis</i>	0.402	0.262	0.248	0.152	0.238	0.246	0.095	0.174	0.306	0.104	0.167	0.141	0.223	0.242	0.168	0.171	0.167	0.160	0.153	0.175	0.143

Table 3. cont.

No.	Species	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
1	<i>P. multocida</i>	0.023	0.059	0.063	0.066	0.059	0.068	0.063	0.063	0.059	0.055	0.068	0.025	0.174	0.066	0.055	0.059	0.070	0.039	0.941	0.941	0.928	0.338
2	<i>H. influenzae</i>	0.059	0.061	0.057	0.059	0.059	0.059	0.057	0.052	0.048	0.045	0.057	0.077	0.300	0.120	0.097	0.084	0.079	0.102	0.203	0.194	0.199	0.226
3	<i>H. somni</i>	0.005	0.025	0.014	0.023	0.021	0.023	0.014	0.014	0.019	0.021	0.019	-0.018	0.194	0.014	0.032	0.019	0.012	0.025	0.176	0.172	0.178	0.160
4	<i>G. anatis</i>	0.037	0.016	0.025	0.025	0.019	0.016	0.010	0.007	0.010	0.007	0.007	-0.051	-0.018	-0.027	-0.024	-0.051	-0.031	-0.035	0.016	0.014	0.014	-0.018
5	<i>V. psittacida</i>	0.086	0.091	0.084	0.093	0.088	0.084	0.081	0.081	0.073	0.075	0.084	0.093	0.203	0.120	0.084	0.111	0.052	0.066	0.226	0.223	0.221	0.183
6	<i>A. gallinarum</i>	0.019	0.039	0.012	0.014	0.027	0.030	0.021	0.021	0.023	0.021	0.021	0.086	0.228	0.099	0.079	0.063	0.066	0.057	0.178	0.174	0.176	0.154
7	<i>N. semolina</i>	0.313	0.329	0.300	0.300	0.316	0.316	0.311	0.309	0.318	0.316	0.313	0.379	0.057	0.410	0.248	0.284	0.277	0.284	0.005	0.007	-0.002	0.030
8	<i>P. uteri</i>	0.246	0.253	0.264	0.266	0.230	0.237	0.232	0.230	0.228	0.228	0.235	0.262	0.075	0.280	0.253	0.232	0.235	0.217	0.059	0.057	0.059	0.021
9	<i>L. koalarum</i>	0.048	0.063	0.050	0.059	0.045	0.050	0.052	0.052	0.045	0.048	0.055	0.041	0.325	0.088	0.075	0.077	0.055	0.037	0.246	0.248	0.255	0.217
10	<i>M. haemolytica</i>	0.266	0.259	0.295	0.295	0.257	0.257	0.253	0.250	0.257	0.255	0.255	0.295	0.099	0.329	0.921	0.471	0.520	0.421	0.052	0.045	0.043	0.007
11	[H.] <i>ducreyi</i>	0.352	0.352	0.340	0.349	0.325	0.327	0.320	0.320	0.325	0.325	0.322	0.268	0.057	0.313	0.273	0.264	0.266	0.246	0.041	0.043	0.043	0.016
12	[P.] <i>trehalosi</i>	0.325	0.338	0.329	0.336	0.293	0.304	0.295	0.295	0.302	0.300	0.298	0.340	0.106	0.374	0.295	0.284	0.300	0.271	0.059	0.061	0.070	0.048
13	[A.] <i>actinomycetemcomitans</i>	0.052	0.023	0.019	0.021	0.010	0.010	0.010	0.007	0.001	0.003	0.007	0.050	0.264	0.075	0.025	0.019	0.021	-0.002	0.192	0.194	0.192	0.167
14	[M.] <i>succiniciproducens</i>	0.057	0.052	0.084	0.086	0.039	0.037	0.034	0.034	0.039	0.034	0.037	0.097	0.284	0.124	0.045	0.019	0.021	0.048	0.185	0.178	0.183	0.131
15	[A.] <i>minor</i>	0.345	0.336	0.329	0.334	0.329	0.329	0.325	0.325	0.311	0.311	0.325	0.334	0.091	0.356	0.313	0.318	0.318	0.311	0.052	0.052	0.059	0.037
16	[A.] <i>porcitonsillarum</i>	0.345	0.329	0.331	0.331	0.334	0.334	0.334	0.334	0.320	0.322	0.334	0.349	0.095	0.358	0.300	0.313	0.327	0.291	0.061	0.066	0.066	0.050
17	<i>Actinobacillus</i> genomospecies 2	0.687	0.653	0.664	0.660	0.619	0.631	0.628	0.628	0.628	0.626	0.628	0.322	0.077	0.358	0.266	0.271	0.284	0.286	0.055	0.059	0.066	0.003
18	<i>A. hominis</i>	0.709	0.689	0.754	0.754	0.644	0.646	0.640	0.635	0.649	0.646	0.635	0.307	0.073	0.343	0.268	0.268	0.282	0.275	0.070	0.066	0.073	0.025
19	<i>A. ureae</i>	0.712	0.694	0.754	0.748	0.644	0.644	0.640	0.637	0.651	0.649	0.637	0.304	0.077	0.343	0.275	0.273	0.289	0.280	0.077	0.075	0.081	0.025
20	<i>A. arthritidis</i>	0.682	0.658	0.662	0.664	0.624	0.633	0.633	0.631	0.631	0.631	0.633	0.316	0.061	0.354	0.255	0.266	0.271	0.280	0.043	0.048	0.057	-0.004
21	<i>A. suis</i>	0.743	0.770	0.741	0.739	0.703	0.716	0.716	0.714	0.748	0.752	0.716	0.340	0.081	0.356	0.250	0.282	0.295	0.280	0.052	0.052	0.063	-0.002
22	<i>Actinobacillus</i> genomospecies 1	0.730	0.768	0.772	0.698	0.709	0.705	0.705	0.730	0.727	0.703	0.325	0.079	0.358	0.264	0.275	0.280	0.259	0.023	0.025	0.032	0.010	
23	<i>A. capsulatus</i>	0.755	0.739	0.734	0.709	0.725	0.725	0.723	0.732	0.732	0.725	0.325	0.077	0.345	0.264	0.282	0.284	0.282	0.057	0.057	0.063	-0.008	
24	<i>A. equuli</i> subsp. <i>equuli</i>	0.767	0.794	0.946	0.678	0.691	0.685	0.682	0.696	0.694	0.685	0.322	0.081	0.343	0.291	0.284	0.298	0.277	0.063	0.063	0.061	0.021	
25	<i>A. equuli</i> subsp. <i>haemolyticus</i>	0.786	0.802	0.872	0.676	0.687	0.682	0.680	0.694	0.694	0.682	0.322	0.084	0.347	0.293	0.286	0.295	0.277	0.066	0.066	0.066	0.023	
26	<i>A. lignieresii</i>	0.745	0.738	0.751	0.734	0.910	0.905	0.903	0.871	0.869	0.903	0.304	0.061	0.343	0.255	0.253	0.293	0.268	0.057	0.052	0.061	-0.029	
27	<i>A. pleuropneumoniae</i> serotype 1	0.745	0.735	0.744	0.730	0.908	0.946	0.943	0.912	0.910	0.943	0.311	0.073	0.345	0.255	0.255	0.298	0.275	0.066	0.063	0.070	-0.029	
28	<i>A. pleuropneumoniae</i> serotype 2	0.745	0.739	0.746	0.733	0.913	0.933	0.946	0.907	0.910	0.946	0.309	0.070	0.347	0.253	0.246	0.293	0.266	0.063	0.057	0.063	-0.029	
29	<i>A. pleuropneumoniae</i> serotype 3	0.744	0.736	0.744	0.730	0.912	0.931	0.937	0.907	0.905	0.948	0.304	0.068	0.345	0.250	0.246	0.291	0.266	0.063	0.057	0.063	-0.031	
30	<i>A. pleuropneumoniae</i> serotype 4	0.751	0.740	0.747	0.735	0.904	0.926	0.930	0.928	0.946	0.907	0.318	0.073	0.356	0.253	0.253	0.298	0.271	0.057	0.052	0.059	-0.031	
31	<i>A. pleuropneumoniae</i> serotype 6	0.751	0.741	0.747	0.736	0.905	0.925	0.930	0.927	0.938	0.905	0.316	0.075	0.354	0.250	0.253	0.295	0.271	0.052	0.045	0.055	-0.029	
32	<i>A. pleuropneumoniae</i> serotype 8	0.743	0.736	0.743	0.730	0.911	0.929	0.934	0.935	0.926	0.925	0.309	0.073	0.347	0.253	0.248	0.295	0.268	0.066	0.061	0.068	-0.029	
33	[A.] <i>indolicus</i>	0.343	0.338	0.329	0.339	0.324	0.326	0.323	0.324	0.328	0.328	0.326	0.091	0.842	0.293	0.316	0.325	0.277	0.023	0.027	0.023	0.012	
34	[A.] <i>porcinus</i>	0.224	0.225	0.213	0.213	0.195	0.202	0.201	0.201	0.202	0.202	0.204	0.213	0.133	0.102	0.093	0.061	0.099	0.176	0.172	0.172	0.183	
35	[H.] <i>parasuis</i>	0.330	0.327	0.318	0.334	0.307	0.311	0.310	0.310	0.314	0.314	0.312	0.596	0.205	0.322	0.322	0.365	0.322	0.068	0.057	0.063	0.032	
36	<i>M. glucosida</i>	0.331	0.339	0.336	0.346	0.321	0.326	0.326	0.325	0.325	0.325	0.325	0.336	0.210	0.323	0.473	0.514	0.415	0.057	0.052	0.048	0.016	
37	<i>M. varigena</i>	0.332	0.330	0.323	0.328	0.333	0.338	0.335	0.335	0.338	0.338	0.337	0.330	0.183	0.330	0.512	0.475	0.446	0.059	0.066	0.066	-0.008	
38	<i>M. ruminalis</i>	0.346	0.340	0.338	0.355	0.330	0.343	0.340	0.340	0.343	0.343	0.343	0.334	0.204	0.341	0.733	0.517	0.403	0.073	0.070	0.073	0.019	
39	<i>M. granulomatis</i>	0.330	0.334	0.341	0.345	0.340	0.346	0.342	0.342	0.345	0.346	0.345	0.327	0.191	0.316	0.477	0.475	0.480	0.039	0.045	0.043	-0.024	
40	<i>P. multocida</i> Pm70	0.143	0.145	0.139	0.139	0.132	0.139	0.135	0.137	0.137	0.136	0.140	0.177	0.258	0.192	0.148	0.143	0.151	0.144	0.937	0.921	0.336	
41	<i>P. multocida</i> subsp. <i>gallicida</i>	0.150	0.159	0.150	0.148	0.142	0.147	0.144	0.146	0.145	0.144	0.147	0.179	0.260	0.194	0.147	0.146	0.161	0.145	0.914	0.925	0.336	
42	<i>P. multocida</i> subsp. <i>septica</i>	0.139	0.154	0.138	0.142	0.135	0.139	0.137	0.135	0.136	0.136	0.139	0.182	0.265	0.194	0.156	0.152	0.167	0.141	0.869	0.879	0.334	
43	<i>P. canis</i>	0.158	0.160	0.153	0.157	0.132	0.134	0.134	0.135	0.133	0.133	0.134	0.156	0.279	0.179	0.113	0.109	0.116	0.112	0.412	0.402	0.404	

Comparison of the conventional and novel methods within *Mannheimia*

In order to compare the sequence-based method using the three genes to predict genome similarity with data from classical DNA–DNA hybridization, we selected the genus *Mannheimia* (Angen *et al.*, 1999), a recently described genus which has been clearly circumscribed based on a polyphasic approach and contains several species. Moreover, most data have been generated in the same laboratory under the same methodological and experimental conditions, thereby giving the most reproducible values and avoiding the inherent variability observed between different methods and laboratories. Table 4 shows the values published and calculated from the sequences. Regression analysis showed a very good correlation between published and calculated values ($r=0.980$). Nevertheless, similarity values obtained with the sequence-based method should not be compared directly with those derived from DNA–DNA hybridization, since each method represents a comparative system of its own. For example, with the hybridization method, *M. haemolytica* showed the same value of 63 % with *M. glucosida* and *Mannheimia ruminantis*, even though the sequence-derived similarity values were 0.88 and 0.74, respectively. All the *Mannheimia* species showed similarity values above the threshold for genus separation of 0.4. [*Pasteurella*] *trehalosi* and *Actinobacillus capsulatus* had similarity values with *M. haemolytica* clearly below this limit. Moreover, based on the values published by Angen *et al.* (1999), *Mannheimia granulomatis* showed 34 % DNA–DNA similarity by hybridization with *M. ruminantis* and *A. capsulatus*, although the sequence similarity values were 0.48 and 0.34, respectively (Table 3, lower left), placing the two *Mannheimia* species in the same genus and clearly separating *A. capsulatus* from the genus *Mannheimia*. According to the present investigation, the recently described [*Mannheimia*] *succiniciproducens* has been misclassified and consequently misnamed. Since it did not show significant similarity values with other genera and also phylogenetically is outside the genus *Mannheimia* (Table 3, lower left), it should be excluded from *Mannheimia sensu stricto*. Looking at the species separation based on predicted genome similarity, only the similarity value of 0.88 that *M. glucosida* shows with *M. haemolytica* needs additional interpretation. The similarity value was between 0.85 and 0.9, which indicates that the two are very closely related, as previously shown by Angen *et al.* (1999), and species separation is not obvious according to the sequence-based method alone. *M. haemolytica* and *M. glucosida* are also phylogenetically very close (Korczak *et al.*, 2004). However, additional strains of other species should be sequenced to investigate more precisely the genomic similarity at the species level in comparison with previous DNA–DNA hybridizations.

Analysis of the genus *Actinobacillus sensu stricto*

To investigate the power of the new approach at genus and species level, the genus *Actinobacillus* was chosen as another

Table 4. Comparison of sequencing method with hybridization method within *Mannheimia* and related species

NA, Not available.

Species	<i>Mannheimia haemolytica</i>	
	<i>recN/rpoA/thdF</i>	Published*
<i>Mannheimia glucosida</i>	0.88	63 ± 5 %
<i>Mannheimia varigena</i>	0.51	34 ± 19 %
<i>Mannheimia ruminantis</i>	0.74	63 ± 3 %
<i>Mannheimia granulomatis</i>	0.48	32 ± 12 %
[<i>Pasteurella</i>] <i>trehalosi</i>	0.32	18 %
<i>Actinobacillus capsulatus</i>	0.32	21 ± 11 %
[<i>Mannheimia</i>] <i>succiniciproducens</i>	0.15	NA

*Angen *et al.* (1999).

example. This genus is less well defined than *Mannheimia*, and it is unclear whether certain species, such as *A. capsulatus*, should be considered to be inside or outside the genus, based mainly on *rrs* sequences (Olsen & Moller, 2005). In addition, the type species of the genus, *Actinobacillus lignieresii*, is almost identical to *A. pleuropneumoniae* (Christensen & Bisgaard, 2004).

Strains belonging to *Actinobacillus sensu stricto* showed very high similarity values when compared with each other (Table 3, lower left). This included *A. capsulatus*, which showed a similarity value of 0.74 with the type species *A. lignieresii*. Based on this figure, *A. capsulatus* should be classified as a true member of *Actinobacillus sensu stricto*. This is further supported by earlier DNA–DNA hybridization studies, as well as *rpoB* and *infB* phylogeny (Mutters *et al.*, 1989; Korczak *et al.*, 2004; Norskov-Lauritsen *et al.*, 2004). Further studies that include more strains than the type strain alone need to be carried out in order to finally classify *A. capsulatus*. [*Actinobacillus*] *minor* and '[*Actinobacillus*] *porcitosillarum*' are closely related to the *Actinobacillus sensu stricto* cluster, though forming a small branch of their own in *rrs* as well as *rpoB* trees (Korczak *et al.*, 2004). In Table 3 (lower left), these two taxa showed similarity values below 0.4 with all members of *Actinobacillus sensu stricto*. However, between [*A.*] *minor* and '[*A.*] *porcitosillarum*', the similarity value was 0.79, which would combine them in their own genus, although separating them as two species.

The other two porcine *Actinobacillus* species [*A.*] *indolicus* and [*Actinobacillus*] *porcinus* included in our study were clearly outside the *Actinobacillus sensu stricto* cluster, and did not show genetic relatedness with each other. However, [*A.*] *indolicus* showed a high similarity value of 0.6 with [*H.*] *parasuis*, indicating that these two species form a genus of their own.

Variation within species

In order to examine intraspecies variation we investigated a few serotypes of *A. pleuropneumoniae*. As seen in Table 3

Table 5. G+C content comparison between calculation based on the three genes and published values

No.	Species	mol% G+C content		
		Calculated	Published	Difference*
1	<i>Pasteurella multocida</i> CCUG 17976 ^T	41.8	40.8–43.9†	0.5
2	<i>Haemophilus influenzae</i> Rd	38.5	39‡	0.5
3	<i>Histophilus somni</i> HS8025 ^T	38.7	37.5§	0.4
4	<i>Gallibacterium anatis</i> CCUG 15563 ^T	41.4	39.9–42.3	0.3
5	<i>Volucibacter psittacida</i> JEO101	39.8	39.9–40.8¶	0.6
6	<i>Avibacterium gallinarum</i> NCTC 11188 ^T	43.3	43.6#	0.2
7	<i>Nicotella semolina</i> CCUG 43639 ^T	42.7	NA	—
8	<i>Phocoenobacter uteri</i> NCTC 12872 ^T	38.4	41.5†	3.1
9	<i>Lonepinella koalarum</i> ATCC 700131 ^T	39.3	37.5†	1.8
10	<i>Mannheimia haemolytica</i> NCTC 9380 ^T	42.1	43.6†	1.5
11	[<i>Haemophilus</i>] <i>ducreyi</i> 35000HP	39.1	38†	1.1
12	[<i>Pasteurella</i>] <i>trehalosi</i> NCTC 10370 ^T	41.8	42.6†	0.8
13	[<i>Actinobacillus</i>] <i>actinomycetemcomitans</i> HK1651	45.7	42.7†	3
14	[<i>Mannheimia</i>] <i>succiniciproducens</i> MBEL55E	43.9	42.5**	1.4
15	[<i>Actinobacillus</i>] <i>minor</i> NM305 ^T	41.2	38.2†	3
16	[<i>Actinobacillus</i>] <i>porcitonsillarum</i> CCUG 46996	41.7	NA	—
17	<i>Actinobacillus</i> genomospecies 2 CCUG 15571	41.3	NA	—
18	<i>Actinobacillus hominis</i> CCUG 19800 ^T	42.6	40.9†	1.7
19	<i>Actinobacillus ureae</i> CCUG 2139 ^T	42.3	41.2–43.7†	0.2
20	<i>Actinobacillus arthritidis</i> CCUG 24862 ^T	41.0	NA	—
21	<i>Actinobacillus suis</i> ATCC 33415 ^T	43.0	40.5†	2.5
22	<i>Actinobacillus</i> genomospecies 1 CCUG 22229	43.1	NA	—
23	<i>Actinobacillus capsulatus</i> CCUG 12396 ^T	42.4	42.4†	0
24	<i>Actinobacillus equuli</i> subsp. <i>equuli</i> ATCC 19392 ^T	42.6	40–43.4††	0.9
25	<i>A. equuli</i> subsp. <i>haemolyticus</i> CCUG 19799 ^T	42.4	40–43.4††	0.7
26	<i>Actinobacillus lignieresii</i> NCTC 4189 ^T	43.5	41.8–42.6†	1.3
27	<i>Actinobacillus pleuropneumoniae</i> serotype 1 S4074 ^T	43.9	42.2–43.2†	1.2
28	<i>A. pleuropneumoniae</i> serotype 2 S1536	43.8	42.2–43.2†	1.1
29	<i>A. pleuropneumoniae</i> serotype 3 S1421	43.8	42.2–43.2†	1.1
30	<i>A. pleuropneumoniae</i> serotype 4 M62	43.8	42.2–43.2†	1.1
31	<i>A. pleuropneumoniae</i> serotype 6 FemØ	43.8	42.2–43.2†	1.1
32	<i>A. pleuropneumoniae</i> serotype 8 405	43.6	42.2–43.2†	0.9
33	[<i>Actinobacillus</i>] <i>indolicus</i> 46KC2 ^T	42.6	35.5†	7.1
34	[<i>Actinobacillus</i>] <i>porcinus</i> NM319 ^T	41.6	41.4†	0.2
35	[<i>Haemophilus</i>] <i>parasuis</i> CCUG 3712 ^T	41.2	41–42†	0.3
36	<i>Mannheimia glucosida</i> CCUG 38457 ^T	42.1	41.6†	0.5
37	<i>Mannheimia varigena</i> CCUG 38462 ^T	41.8	41.7†	0.1
38	<i>Mannheimia ruminalis</i> CCUG 38470 ^T	42.0	NA	—
39	<i>Mannheimia granulomatis</i> ATCC 49244 ^T	41.5	39.2†	2.3
40	<i>Pasteurella multocida</i> subsp. <i>multocida</i> Pm70	41.5	40‡‡	1.5
41	<i>P. multocida</i> subsp. <i>gallicida</i> CCUG 17978 ^T	41.8	41.2–42.5†	0
42	<i>P. multocida</i> subsp. <i>septica</i> CCUG 17977 ^T	42.3	41.5–43.5†	0.2
43	<i>Pasteurella canis</i> CCUG 12400 ^T	38.4	37.7–39.8†	0.3

*Mean values were used for calculation.

References: †Olsen *et al.* (2005); ‡Fleischmann *et al.* (1995); §Angen *et al.* (2003); ||Christensen *et al.* (2003a); ¶Christensen *et al.* (2004a); #Blackall *et al.* (2005); **Hong *et al.* (2004); ††Christensen *et al.* (2002); ‡‡May *et al.* (2001).

(lower left), the obtained similarity values were clearly above 0.9. *A. pleuropneumoniae* and *A. lignieresii* also showed a similarity value above 0.9 and could be classified as subspecies of *A. lignieresii*, corresponding to major

differences in disease manifestations. Intraspecies variability can also be seen with *P. multocida* and its subspecies. Whereas *P. multocida* subsp. *gallicida* showed an even higher similarity value to the type species *P. multocida* subsp.

multocida than strain Pm70, *P. multocida* subsp. *septica* had a similarity value below 0.9 but above 0.85, which leaves it open whether it is a subspecies or a species of its own, as argued by several recent papers (Kuhnert *et al.*, 2000; Davies, 2004).

recN alone

Sequence determination of all three genes might be cumbersome. For that reason we investigated whether the *recN* sequence alone might be representative of all three genes, as also proposed by Zeigler (2003). The similarity values are shown in the upper-right panel of Table 3. Negative values were obtained in this approach. This mathematical artifact has no further meaning and only indicates a rather low genetic relatedness between the corresponding taxa. When performing a regression analysis of all similarity values calculated using the three genes versus the values calculated using *recN* alone we found a very good correlation ($r=0.978$). Threshold similarity values for genus and species boundaries are comparable, with very few discrepancies (Fig. 1). Therefore, *recN* could be used as an initial means of comparing genomes of the *Pasteurellaceae* although, as demonstrated by Zeigler (2003), with a lower confidence. For identifying homogeneous species or separation of distantly related organisms it might be sufficient, but for better resolution, the sequence of all three genes would be necessary.

mol% G+C content

The mol% G+C content should ideally be part of the description of a new taxon (Stackebrandt *et al.*, 2002). Again, methods to determine this value can be different (e.g. denaturation temperature or HPLC), and obtained values may differ significantly not only between strains but also between experiments and laboratories. Therefore, published G+C content within a single species varies by up to 5 % for biological as well as methodological reasons (Goodfellow *et al.*, 1997; Olsen *et al.*, 2005). We calculated and summed up the G+C content of the three genes *rpoA*, *recN* and *thdF* and compared it to available published values (Table 5). A mean difference of 1.2 % was observed between published and calculated G+C contents. The difference between calculated and published values did not exceed 3 %, thereby being within the range of variation seen with conventional methods. The extremely high difference observed with [*A.*] *indolicus* (more than 7 %) was surprising (Moller *et al.*, 1996). It should be noted that the published G+C content of only 35.5 % is very low compared to the other members of the family, and especially compared to [*H.*] *parasuis* (41–42 %), to which [*A.*] *indolicus* is most closely related. Therefore, since the G+C content has until now only been determined from the type strain, this difference should be questioned. Interestingly, in the comparison with published genome sequences we saw that the calculated values were generally slightly (about 1 %) higher. One could therefore argue that a correction of –1 % should be considered. Employing housekeeping genes for G+C content calculation

Fig. 1. Correlation of whole-genome similarities calculated by *recN* alone or by including all three genes *recN/rpoA/thdF*. The linear correlation coefficient (r) was 0.978. Circles indicate possible genus and species boundaries at similarity values of around 0.4 and 0.85, respectively.

may reflect the true characteristics of the core genome of a species more accurately, since it is not affected by the variability that might result from horizontally transferred genomic elements, which often show variation from the overall G+C content. In addition to the data on whole-genome similarity, the G+C content calculation showed that the three genes *rpoA*, *recN* and *thdF* can be regarded as ‘genome-representing genes’. The other protein-encoding genes *rpoB* and *infB*, alone or in combination, could not be used for this purpose (data not shown).

Phylogeny based on six housekeeping genes

We have previously used the *rrs*, *rpoB* and *infB* genes for phylogenetic analysis within the family *Pasteurellaceae* (Kuhnert *et al.*, 2004; Korczak *et al.*, 2004). Combining these with the three additional genes we investigated the analysed strains in a multilocus sequence phylogeny (MLSP). Individual phylogenetic trees of each gene were constructed and are available as supplementary data. From the six trees, a consensus tree was generated in Bionumerics (Fig. 2).

The clusters observed corresponded well with the current genera and indicated putative new genera that will have to be described in the future. The clusters also supported the DNA–DNA similarity values obtained from the sequencing method. For example, the species [*M.*] *succiniciproducens* was clearly outside the *Mannheimia* cluster. This is underlined by the very low similarity value of 0.15 of this species to the type species *M. haemolytica* (Table 3, lower left, and Table 4). Within the *Mannheimia* cluster *M. granulomatis* showed the lowest similarity value (0.48) with the type species. This is in agreement with the position in the tree. However, the two species *M. haemolytica* and *M. glucosida*, which showed a high similarity value of 0.88, clustered very closely, like the subspecies of *A. equuli* or *P. multocida*. The

Fig. 2. Consensus tree of six genes. Individual trees of *rrs*, *rpoB*, *infB*, *rpoA*, *recN* and *thdF* were used to generate a combined tree built in Bionumerics. Jukes–Cantor correction was applied for the distance matrix and neighbour-joining for tree construction. Cophenetic correlations are given, indicating the reliability of the branching compared to the actual genetic relatedness of the taxa.

same was observed for *A. linieresii* and *A. pleuropneumoniae*, which also had a very high similarity value above the upper species threshold of 0.9. Moreover, the phylogenetic analysis using all six genes clustered *A. capsulatus* clearly inside *Actinobacillus sensu stricto*. On the other hand, [*A.*] *minor* and [*A.*] *porcitonsillarum* were outside this cluster,

separated from it by [*Haemophilus*] *ducreyi*, which could already be assumed based on the low similarity values that these two species showed with the *sensu stricto* species (Table 3, lower left). Based not only on the similarity values but also on the relatively deep branching of the two, they should be classified as two different species, eventually

Fig. 3. Congruence of trees. The matrix of congruence values (Pearson correlation coefficient) between the genetic distances derived from the six genes as well as the consensus tree (MLSP) is given. A dendrogram derived from this matrix is shown on the left.

forming their own genus. The analysis of more strains and species is necessary in order to reach a final classification of [*A.*] *minor* and [*A.*] *porcitonsillarum*'. The tree also represented the close relatedness of [*A.*] *indolicus* and [*H.*] *parasuis*, supporting the observation from the similarity analysis which indicated that they belong to the same genus. The very deep branching of *Pasteurella canis* in relation to *P. multocida* is also worth mentioning, since this species showed a similarity value of 0.41, which is at the borderline for genus separation. The position of *G. anatis* was the most isolated in the tree, which is again reflected by the similarity values observed for this species with all the others analysed in our study. It is the species that showed the lowest similarity values, ranging from 0.06 to 0.17 (Table 3, lower left). All the other species had significantly higher similarity values on average. Therefore, the phylogenetic tree based on six genes and the similarity values calculated are in very good agreement.

The congruence between the individual trees (see supplementary data) was calculated by Pearson correlation in Bionumerics. Results are presented graphically as well as in tabular form in Fig. 3. From this it can be deduced that *recN* most properly represents the consensus tree (MLSP), indicated by the high correlation between the two (97.1 %). The other genes, *rpoB*, *infB*, *rpoA* and *thdF*, had lower values between 92 and 95 % with the consensus tree. Overall, the three genes *recN*, *rpoA* and *thdF* had tree topologies that were closest to that of the consensus tree. This is further evidence that these genes are highly representative for the entire genome. The genes *rpoB* and *infB* shared some common tree topology, as shown in Fig. 3. The most divergent tree, showing the lowest correlations to all the other trees, was the one derived from *rrs*. This might be explained by the different nature of this gene, since it is not protein coding. Nevertheless, this finding is also surprising, since *rrs* is regarded as the gold standard for phylogenetic analysis of bacteria. However, at least within the *Pasteurellaceae*, it seems as if this gene is less representative than any of

the others used in our study. This has previously been observed within the family *Pasteurellaceae* as well as the genus *Campylobacter* from the fact that *rpoB*-based phylogeny is in better agreement with the results of DNA–DNA hybridization than *rrs*-based phylogeny (Korczak *et al.*, 2004, 2006). Therefore, the *rrs* gene might be suitable for analysing distantly related taxa, but genes other than *rrs* might prove to be more informative for phylogeny and taxonomy in other bacterial families and genera.

Conclusion

From the data presented we conclude that MLSA using the three genes *recN*, *rpoA* and *thdF* represents a valid approach for investigating the taxonomy of *Pasteurellaceae*. The obtained results can be used to calculate whole-genome similarity, they give a very good estimate of the G+C content of a species, and, finally, the genes can be used alone or in combination as phylogenetic markers, allowing a comprehensive analysis of the phylogenetic position of a strain. The results obtained by the sequence-based method are congruent with data from conventional DNA–DNA hybridization at the species level with respect to whole-genome similarity as well as G+C content. We therefore propose to add the MLSA-based approach to those previously used, as recommended by leading taxonomists (Stackebrandt *et al.*, 2002; Gevers *et al.*, 2005). Complete matrix hybridizations between a new putative species and validated species and genera are no longer needed. A simple sequence analysis of selected strains of a new species and calculation of similarity with already-sequenced taxa is enough, and data are highly comparable between laboratories. This would also favour the creation of databases for this purpose, which will allow taxonomists to compare their new taxa at a remote location via the internet. The exchange of reference strains will no longer be necessary, and the new approach could boost the reorganization of the family *Pasteurellaceae*, which is important in the light of many misclassified species and unnamed taxa. Finally, the universal validity of the approach described should be tested with other bacterial families.

ACKNOWLEDGEMENTS

We thank Magne Bisgaard for critical reading of the manuscript and valuable discussions. This work was supported by a grant from the Innovation Promotion Agency of Switzerland (KTI) (no. 6041.1 KTS).

REFERENCES

- Angen, O., Muters, R., Caugant, D. A., Olsen, J. E. & Bisgaard, M. (1999). Taxonomic relationships of the [*Pasteurella*] *haemolytica* complex as evaluated by DNA–DNA hybridizations and 16S rRNA sequencing with proposal of *Mannheimia haemolytica* gen. nov., comb. nov., *Mannheimia granulomatis* comb. nov., *Mannheimia glucosida* sp. nov., *Mannheimia ruminantis* sp. nov. and *Mannheimia varigena* sp. nov. *Int J Syst Bacteriol* **49**, 67–86.
- Angen, O., Ahrens, P., Kuhnert, P., Christensen, H. & Muters, R. (2003). Proposal of *Histophilus somni* gen. nov., sp. nov. for the three

species incertae sedis 'Haemophilus somnus', 'Haemophilus agni' and 'Histophilus ovis'. *Int J Syst Evol Microbiol* **53**, 1449–1456.

Blackall, P. J., Christensen, H., Beckenham, T., Blackall, L. L. & Bisgaard, M. (2005). Reclassification of *Pasteurella gallinarum*, [*Haemophilus*] *paragallinarum*, *Pasteurella avium* and *Pasteurella volantium* as *Avibacterium gallinarum* gen. nov., comb. nov., *Avibacterium paragallinarum* comb. nov., *Avibacterium avium* comb. nov. and *Avibacterium volantium* comb. nov. *Int J Syst Evol Microbiol* **55**, 353–362.

Brenner, D. J., McWhorter, A. C., Knutson, J. K. & Steigerwalt, A. G. (1982). *Escherichia vulneris*: a new species of *Enterobacteriaceae* associated with human wounds. *J Clin Microbiol* **15**, 1133–1140.

Christensen, H. & Bisgaard, M. (2004). Revised definition of *Actinobacillus sensu stricto* isolated from animals. A review with special emphasis on diagnosis. *Vet Microbiol* **99**, 13–30.

Christensen, H., Bisgaard, M. & Olsen, J. E. (2002). Reclassification of equine isolates previously reported as *Actinobacillus equuli*, variants of *A. equuli*, *A. suis* or taxon 11 of Bisgaard and proposal of two subspecies of *A. equuli*: *A. equuli* subsp. *equuli* and *A. equuli* subsp. *haemolyticus*. *Int J Syst Evol Microbiol* **52**, 1569–1576.

Christensen, H., Bisgaard, M., Bojesen, A. M., Mutters, R. & Olsen, J. E. (2003a). Genetic relationships among avian isolates classified as *Pasteurella haemolytica*, '*Actinobacillus salpingitidis*' or *Pasteurella anatis* with proposal of *Gallibacterium anatis* gen. nov., comb. nov. and description of additional genomospecies within *Gallibacterium* gen. nov. *Int J Syst Evol Microbiol* **53**, 275–287.

Christensen, H., Foster, G., Christensen, J. P., Pennycott, T., Olsen, J. E. & Bisgaard, M. (2003b). Phylogenetic analysis by 16S rDNA gene sequence comparison of avian taxa of Bisgaard and characterization and description of two new taxa of *Pasteurellaceae*. *J Appl Microbiol* **95**, 354–363.

Christensen, H., Bisgaard, M., Aalbaek, B. & Olsen, J. E. (2004a). Reclassification of Bisgaard taxon 33 with proposal of *Volucrobacter psittacidica* gen. nov., sp. nov. and *Volucrobacter amazonae* sp. nov. as new members of *Pasteurellaceae*. *Int J Syst Evol Microbiol* **54**, 813–818.

Christensen, H., Kuhnert, P., Olsen, J. E. & Bisgaard, M. (2004b). Comparative phylogenies of the housekeeping genes *atpD*, *infB* and *rpoB* and the 16S rRNA gene within the *Pasteurellaceae*. *Int J Syst Evol Microbiol* **54**, 1601–1609.

Davies, R. L. (2004). Genetic diversity among *Pasteurella multocida* strains of avian, bovine, ovine and porcine origin from England and Wales by comparative sequence analysis of the 16S rRNA gene. *Microbiology* **150**, 4199–4210.

Donachie, W., Lainson, F. A. & Hodgson, J. C. (1995). *Haemophilus*, *Actinobacillus*, and *Pasteurella*. New York: Plenum.

Fleischmann, R. D., Adams, M. D., White, O. & 7 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Gautier, A. L., Dubois, D., Escande, F., Avril, J. L., Trieu-Cuot, P. & Gailliot, O. (2005). Rapid and accurate identification of human isolates of *Pasteurella* and related species by sequencing the *sodA* gene. *J Clin Microbiol* **43**, 2307–2314.

Gevers, D., Cohan, F. M., Lawrence, J. G. & 8 other authors (2005). Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol* **3**, 733–739.

Goodfellow, M., Manfio, G. P. & Chun, J. (1997). Towards a practical species concept for cultivable bacteria. In *Species: the Units of Biodiversity*, pp. 25–59. Edited by M. F. Claridge, H. A. Dawah & M. R. Wilson. London: Chapman and Hall.

Hedegaard, J., Okkels, H., Bruun, B., Kilian, M., Mortensen, K. K. & Nørskov-Lauritsen, N. (2001). Phylogeny of the genus *Haemophilus*

as determined by comparison of partial *infB* sequences. *Microbiology* **147**, 2599–2609.

Hong, S. H., Kim, J. S., Lee, S. Y. & 7 other authors (2004). The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat Biotechnol* **22**, 1275–1281.

Korczak, B., Christensen, H., Emler, S., Frey, J. & Kuhnert, P. (2004). Phylogeny of the family *Pasteurellaceae* based on *rpoB* sequences. *Int J Syst Evol Microbiol* **54**, 1393–1399.

Korczak, B. M., Stieber, R., Emler, S., Burnens, A. P., Frey, J. & Kuhnert, P. (2006). Genetic relatedness within the genus *Campylobacter* inferred from *rpoB* sequences. *Int J Syst Evol Microbiol* **56**, 937–945.

Kuhnert, P., Boerlin, P., Emler, S., Krawinkler, M. & Frey, J. (2000). Phylogenetic analysis of *Pasteurella multocida* subspecies and molecular identification of feline *P. multocida* subsp. *septica* by 16S rRNA gene sequencing. *Int J Med Microbiol* **290**, 599–604.

Kuhnert, P., Frey, J., Lang, N. P. & Mayfield, L. (2002). A phylogenetic analysis of *Prevotella nigrescens*, *Prevotella intermedia* and *Porphyromonas gingivalis* field strains reveals a clear species clustering. *Int J Syst Evol Microbiol* **52**, 1391–1395.

Kuhnert, P., Korczak, B., Falsen, E., Straub, R., Hoops, A., Boerlin, P., Frey, J. & Mutters, R. (2004). *Nicoletella semolina* gen. nov., sp. nov., a new member of *Pasteurellaceae* isolated from horses with airway disease. *J Clin Microbiol* **42**, 5542–5548.

May, B. J., Zhang, Q., Li, L. L., Paustian, M. L., Whittam, T. S. & Kapur, V. (2001). Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc Natl Acad Sci U S A* **98**, 3460–3465.

Møller, K., Fussing, V., Grimont, P. A. D., Paster, B. J., Dewhirst, F. E. & Kilian, M. (1996). *Actinobacillus minor* sp. nov., *Actinobacillus porcicus* sp. nov., and *Actinobacillus indolicus* sp. nov., three new V factor-dependent species from the respiratory tract of pigs. *Int J Syst Bacteriol* **46**, 951–956.

Mutters, R., Mannheim, W. & Bisgaard, M. (1989). Taxonomy of the group. In *Pasteurella and Pasteurellosis*, pp. 3–34. Edited by C. Adlam & J. M. Rutter. London: Academic Press.

Nørskov-Lauritsen, N., Christensen, H., Okkels, H., Kilian, M. & Bruun, B. (2004). Delineation of the genus *Actinobacillus* by comparison of partial *infB* sequences. *Int J Syst Evol Microbiol* **54**, 635–644.

Olsen, I., Dewhirst, F. E., Paster, B. J. & Busse, H. J. (2005). Family *Pasteurellaceae*. In *Bergey's Manual of Systematic Bacteriology*. Edited by R. Garrity. New York: Springer.

Olsen, I. & Møller, K. (2005). Genus *Actinobacillus* Brumpt 1919, 849^{AL}. In *Bergey's Manual of Systematic Bacteriology*. Edited by R. Garrity. New York: Springer.

Pohl, S. (1981). DNA relatedness among members of *Haemophilus*, *Pasteurella* and *Actinobacillus*. In *Haemophilus*, *Pasteurella* and *Actinobacillus*, p. 253. Edited by M. Kilian, W. Frederiksen & E. L. Biberstein. London: Academic Press.

Stackebrandt, E., Frederiksen, W., Garrity, G. M. & 10 other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Wayne, L. G., Brenner, D. J., Colwell, R. R. & 9 other authors (1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**, 463–464.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int J Syst Evol Microbiol* **53**, 1893–1900.